PCT WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 : G10L 5/00

A1

(11) Internationale Veröffentlichungsnummer: **WO 99/40568**

(43) Internationales Veröffentlichungsdatum: 12. August 1999 (12.08.99)

(21) Internationales Aktenzeichen: PCT/DE98/03812

(22) Internationales Anmeldedatum: 30. Dezember 1998 (30.12.98)

(30) Prioritätsdaten:
198 04 189.6 3. Februar 1998 (03.02.98) DE

(71) Anmelder *(für alle Bestimmungsstaaten ausser US)*: SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, D–80333 München (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: HÜNLICH, Klaus [DE/DE]; Birkenstrasse 4, D–85467 Neuching (DE). FRAAS, Wolfgang [DE/DE]; Karwendelstrasse 2, D–82515 Wolfratshausen (DE).

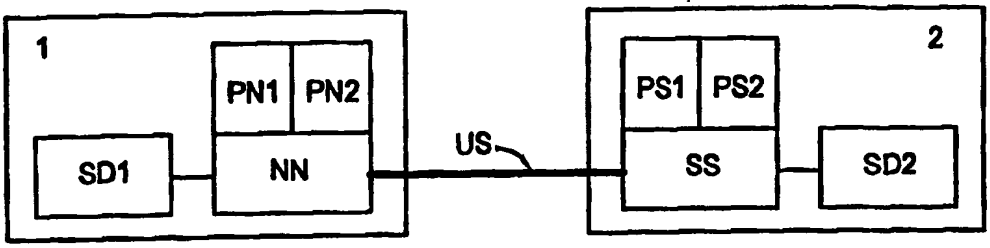
(74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, D–80506 München (DE).

(81) Bestimmungsstaaten: CN, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht
Mit internationalem Recherchenbericht.

(54) Title: METHOD FOR VOICE DATA TRANSMISSION

(54) Bezeichnung: VERFAHREN ZUM ÜBERMITTELN VON SPRACHDATEN



(57) Abstract

In order to transmit voice data, the voice data flow is broken down into phonemes. A code character is assigned to each phoneme in a selective language and/or speaker–specific phoneme catalog (PN1, PN2) and transmitted to a voice synthesis device (SS) located at the transmission target (SD2), whereby the amount of data to be transmitted is considerably reduced. The voice data flow is broken down into phonemes by a neuronal network (NN), which is trained to recognize the phonemes stored in the selective language and/or speaker–specific phoneme catalog (PN1, PN2). The flow of code characters received is once again converted into a sequence of phonemes and emitted by the voice synthesis device (SS).

(57) Zusammenfassung

Zum Übermitteln von Sprachdaten wird der Sprachdatenstrom in Phoneme zerlegt und für jedes Phonem ein diesem in einem auswählbaren sprach– und/oder sprecherspezifischen Phonemkatalog (PN1, PN2) zugeordnetes Kodezeichen zu einer Sprachsyntheseeinrichtung (SS) am Übermittlungsziel (SD2) übertragen, wodurch die zu übertragende Datenmenge i.a. stark reduziert wird. Das Zerlegen des Sprachdatenstroms in Phoneme wird von einem neuronalen Netz (NN) ausgeführt, das auf ein Erkennen der im ausgewählten sprach– und/oder sprecherspezifischen Phonemkatalog (PN1, PN2) hinterlegten Phoneme trainiert ist. Von der Sprachsyntheseeinrichtung (SS) wird der Strom empfangener Kodezeichen wieder in eine Abfolge von Phonemen umgesetzt und ausgegeben.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	**ES**	Spanien	**LS**	Lesotho	**SI**	Slowenien
AM	Armenien	**FI**	Finnland	**LT**	Litauen	**SK**	Slowakei
AT	Österreich	**FR**	Frankreich	**LU**	Luxemburg	**SN**	Senegal
AU	Australien	**GA**	Gabun	**LV**	Lettland	**SZ**	Swasiland
AZ	Aserbaidschan	**GB**	Vereinigtes Königreich	**MC**	Monaco	**TD**	Tschad
BA	Bosnien-Herzegowina	**GE**	Georgien	**MD**	Republik Moldau	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagaskar	**TJ**	Tadschikistan
BE	Belgien	**GN**	Guinea	**MK**	Die ehemalige jugoslawische Republik Mazedonien	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Griechenland			**TR**	Türkei
BG	Bulgarien	**HU**	Ungarn	**ML**	Mali	**TT**	Trinidad und Tobago
BJ	Benin	**IE**	Irland	**MN**	Mongolei	**UA**	Ukraine
BR	Brasilien	**IL**	Israel	**MR**	Mauretanien	**UG**	Uganda
BY	Belarus	**IS**	Island	**MW**	Malawi	**US**	Vereinigte Staaten von Amerika
CA	Kanada	**IT**	Italien	**MX**	Mexiko		
CF	Zentralafrikanische Republik	**JP**	Japan	**NE**	Niger	**UZ**	Usbekistan
CG	Kongo	**KE**	Kenia	**NL**	Niederlande	**VN**	Vietnam
CH	Schweiz	**KG**	Kirgisistan	**NO**	Norwegen	**YU**	Jugoslawien
CI	Côte d'Ivoire	**KP**	Demokratische Volksrepublik Korea	**NZ**	Neuseeland	**ZW**	Zimbabwe
CM	Kamerun			**PL**	Polen		
CN	China	**KR**	Republik Korea	**PT**	Portugal		
CU	Kuba	**KZ**	Kasachstan	**RO**	Rumänien		
CZ	Tschechische Republik	**LC**	St. Lucia	**RU**	Russische Föderation		
DE	Deutschland	**LI**	Liechtenstein	**SD**	Sudan		
DK	Dänemark	**LK**	Sri Lanka	**SE**	Schweden		
EE	Estland	**LR**	Liberia	**SG**	Singapur		

Beschreibung

Verfahren zum Übermitteln von Sprachdaten

Die vorliegende Erfindung betrifft ein Verfahren zum Übermitteln von Sprachdaten, wobei die Sprachdaten vor dem Übermitteln komprimiert und am Übertragungsziel wieder dekomprimiert werden. Das Komprimieren beruht dabei auf einem Zerlegen der Sprachdaten in Phoneme. Phoneme sind die für die Wahrnehmung gesprochener Sprache wesentlichen akustischen Sprachelemente.

Es ist bekannt, Sprachdaten vor einem Übermitteln in einem Kommunikationsnetz zu komprimieren, um möglichst wenig Übertragungsbandbreite im Kommunikationsnetz zu belegen. Bei einer Wiedergabe der Sprache am Übermittlungsziel werden in diesen Fällen die komprimierten Sprachdaten durch Dekomprimieren wieder in den Ursprungszustand oder einen dazu gleichwertigen Zustand zurückgeführt. Da sich die durch ein solches Verfahren erzielbare Reduktion der Übertragungsbandbreite direkt nach der Kompressionsrate des verwendeten Komprimierungsverfahrens richtet, ist eine möglichst hohe Kompressionsrate erstrebenswert.

Bei Sprachübermittlung werden üblicherweise Prädiktionsverfahren zum Komprimieren eingesetzt, welche die statistische Ungleichverteilung der in Sprachdaten auftretenden Datenmuster ausnutzen, um eine Sprachdaten eigentümliche, hohe Redundanz zu reduzieren. Beim Dekomprimieren können die ursprünglichen Sprachdaten aus den komprimierten Sprachdaten bis auf geringe, verfahrensinhärente Verluste, nahezu verfälschungsfrei rekonstruiert werden. Die erreichbare Kompressionsrate liegt dabei in der Größenordnung von etwa 1:10. Verfahren dieser Art sind beispielsweise in „Information und Codierung" von Richard W. Hamming, VCH Verlagsgesellschaft Weinheim, 1987, Seiten 81 bis 97 beschrieben.

In typischen Sprachdaten macht eine rein inhaltliche Information nur einen geringen Bruchteil der gesamten Sprachinformation aus. Der größte Teil der Sprachinformation besteht in der Regel aus sprecherspezifischer Information, die sich z.B. in Nuancen der Stimme oder der Stimmlage des Sprechers ausdrückt. Kommt es bei einer Übermittlung von Sprachdaten im wesentlichen nur auf deren inhaltliche Information an, z.B. bei rein informativen Mitteilungen, automatischen Ansagen o.ä., sind daher durch Reduktion der sprecherspezifischen Information auch wesentlich höhere Kompressionsraten möglich als mit Verfahren, die den Informationsgehalt der Sprachdaten vollständig oder nahezu vollständig erhalten.

Die kleinsten akustischen Einheiten, in denen Sprache vom Sprecher formuliert wird und in denen sich auch die inhaltliche Information - die gesprochenen Wörter - ausdrückt, sind Phoneme. Aus den Patentschriften EP 71716 B1, DE 3513243 C2 und EP 423800 B1 sind bereits Anordnungen und Verfahren bekannt, bei denen ein Sprachdatenstrom hinsichtlich der darin enthaltenen Phoneme analysiert und in einen Strom von den erkannten Phonemen jeweils zugeordneten Kodezeichen umgesetzt wird, um die Sprachdaten dadurch vor einem Übermitteln zu komprimieren.

Ein wesentliches Problem besteht dabei darin, die Phoneme, aus welchen sich ein beliebiger zu übermittelnder Sprachdatenstrom zusammensetzt, zuverlässig zu erkennen. Dies wird insbesondere dadurch erschwert, daß dasselbe Phonem je nach Sprecher und dessen Sprachgewohnheit in sehr unterschiedlicher Ausprägung auftreten kann. Werden Phoneme innerhalb des Sprachdatenstroms nicht erkannt oder falschen Lauten zugeordnet, verschlechtert sich die Übertragungsqualität der Sprache - unter Umständen bis zur Unverständlichkeit. Eine verläßliche Phonemanalyse ist daher ein wichtiges Kriterium für die Qualität bzw. die Anwendungsbreite derartiger Sprachübermittlungsverfahren.

Aufgabe der vorliegenden Erfindung ist es, ein flexibles und effizientes Verfahren zum Übermitteln von Sprachdaten anzugeben, bei dem Sprachdaten durch eine verbesserte Phonemanalyse vor dem Übermitteln komprimiert werden können.

Gelöst wird diese Aufgabe erfindungsgemäß durch ein Verfahren mit den Merkmalen des Patentanspruchs 1.

Vorteilhafte Weiterbildungen der Erfindung sind in den Unteransprüchen angegeben.

Bei einem erfindungsgemäßen Verfahren werden von einer Sprachdatenquelle zu einem Übermittlungsziel zu übermittelnde Sprachdaten vor dem eigentlichen Übertragen einer Phonemanalyse unterzogen. Zur Anwendung des Verfahrens können die Sprachdaten in vielerlei Form vorliegen; beispielsweise in analoger oder digitalisierter Form oder als Sprachsignale beschreibende Merkmalsvektoren, jeweils in zeit- und/oder frequenzaufgelöster Darstellung. Die Phonemanalyse wird erfindungsgemäß durch ein neuronales Netz ausgeführt, das auf ein Erkennen von Phonemen trainiert ist. Die Prinzipien einer Sprach- bzw. Phonemerkennung mittels neuronaler Netze sind beispielsweise in „Review of Neutral Networks for Speech Recognition" von R. P. Lippmann in Neural Computation 1, 1989, Seiten 1 bis 38 beschrieben.

Die Phoneme nach denen der Sprachdatenstrom zu analysieren ist und auf die das neuronale Netz trainiert ist, sind in sprach- und/oder sprecherspezifischen Phonemkatalogen hinterlegt, wo ihnen jeweils ein eindeutiges Kodezeichen, z.B. ein Index oder eine Nummer, zugeordnet ist. Als Sprache können in diesem Zusammenhang u.a. Landessprachen, regionale Dialekte und Ausprägungen von Sprachbesonderheiten verstanden werden. Die sprach- bzw. sprecherspezifischen Phonemkataloge können in einem das erfindungsgemäße Verfahren vorbereitenden Schritt durch Analyse von jeweils unterschiedlichen Sprachen angehörenden und/oder von unterschiedlichen Sprechern stam-

menden, repräsentativen Sprachdaten erstellt werden. Dabei werden diejenigen akustischen Sprachelemente, die sich für die jeweiligen repräsentativen Sprachdaten als charakteristisch erweisen oder eine repräsentative Idealform dieser Sprachelemente, als Phoneme im betreffenden Phonemkatalog hinterlegt. Die Analyse der repräsentativen Sprachdaten kann insbesondere auch mit Hilfe eines neuronalen Netzes durchgeführt werden. Ferner kann die Auswahl von repräsentativen Sprachdaten bzw. deren charakteristischen Sprachelementen vorteilhafterweise Besonderheiten der zu übermittelnden Sprachdaten, wie z.B. Hintergrundgeräuschen o.ä., angepaßt werden.

Ein Phonemkatalog kann beispielsweise als ein konventioneller Speicher und/oder als Teil eines trainierten neuronalen Netzes realisiert sein, wobei im letzten Fall die hinterlegten Phoneme durch einen inneren Zustand des trainierten neuronalen Netzes repräsentiert sein können. Weiterhin kann die den Lernzustand des neuronalen Netzes bestimmende, sogenannte Gewichtsmatrix als Phonemkatalog aufgefaßt werden. Eine solche Gewichtsmatrix kann aus einem trainierten neuronalen Netz ausgelesen, in einem Speicher hinterlegt und bei Bedarf in ein beliebiges neuronales Netz eingelesen werden, das dadurch zu einem entsprechend trainierten neuronalen Netz wird.

Vor der Phonemanalyse der Sprachdaten durch ein neuronales Netz wird zunächst einer der sprach- und/oder sprecherspezifischen Phomenkataloge ausgewählt und die Sprachdaten anschließend dem auf das Erkennen der in dem ausgewählten Phonemkatalog hinterlegten Phoneme trainierten neuronalen Netz zur Phonemanalyse zugeführt. Die Auswahl des Phonemkatalogs kann dabei abhängig von einer Kennung eines die zu übertragenden Sprachdaten sendenden Teilnehmers erfolgen oder von einem sprach- und/oder sprecherunterscheidenden neuronalen Netz veranlaßt werden, das auf ein Erkennen der Sprache, der die zu übertragenden Sprachdaten angehören, und/oder des

Sprechers, von dem die zu übertragenden Sprachdaten stammen, trainiert ist.

Für die bei der Phonemanalyse in den Sprachdaten erkannten Phoneme wird das diesen im ausgewählten Phonemkatalog jeweils zugeordnete Kodezeichen ermittelt und anstelle des dem jeweiligen Phonem entsprechenden Sprachdatenmusters zum Übermittlungsziel übertragen. Da die übertragenen Kodezeichen im allgemeinen wesentlich kürzer als die dadurch repräsentierten Sprachdatenmuster sind, wird die zu übertragende Datenmenge im allgemeinen wesentlich reduziert und so eine hohe Kompressionsrate erzielt. Am Übermittlungsziel werden die empfangenen Kodezeichen einer Sprachsyntheseeinrichtung zugeführt, die den Strom der Kodezeichen wieder in eine Abfolge von Phonemen umsetzt und ausgibt. Zum Ersetzen der empfangenen Kodezeichen durch auszugebende Phoneme greift die Sprachsyntheseeinrichtung dabei ebenfalls auf einen Phonemkatalog zu. Die Kompressionsrate kann eventuell noch gesteigert werden, indem der Strom der Kodezeichen vor dessen Übertragung durch Anwendung eines weiteren Komprimierungsverfahrens komprimiert und am Übermittlungsziel vor dem Zuführen zur Sprachsyntheseeinrichtung wieder dekomprimiert wird.

Die Verwendung sprecher- bzw. sprachspezifischer Phonemkataloge erhöht die Zuverlässigkeit der Phonemanalyse und damit auch die Qualität der Sprachsynthese insbesondere bei Übertragung von Sprachdaten, die von verschiedenen Sprechern stammen. Weiterhin kann durch die Wahl des der Sprachsynthese zugrunde liegenden Phonemkatalogs die Wiedergabe der Sprachdaten auch vorgebbaren Erfordernissen angepaßt werden, z.B. um bestimmte Sprecher oder Sprachbesonderheiten zu simulieren oder um durch ein Verändern der Stimme auf besondere Umstände aufmerksam zu machen.

Nach einer vorteilhaften Weiterbindung des erfindugsgemäßen Verfahrens kann ein Phonemkatalog während einer Übermittlung von Sprachdaten durch Analyse dieser konkret zu übermitteln-

den Sprachdaten - gegebenenfalls mehrmals - aktualisiert werden und/oder das neuronale Netz mittels der zu übermittelnden Sprachdaten weiter trainiert werden. Auf diese Weise kann die Phonemanalyse bzw. die Sprachsynthese auch kurzfristigen Änderungen von Spracheigenheiten angepaßt werden.

Gemäß einer Weiterbildung der Erfindung können Informationen über einen bei der Phonemanalyse verwendeten Phonemkatalog der Sprachsyntheseeinrichtung übermittelt werden. Auf diese Weise ist es beispielsweise möglich, der Sprachsynteseeinrichtung einen Wechsel eines Phonemkataloges, z.B. aufgrund eines Sprecher- oder Sprachwechsels, anzuzeigen oder, falls ein zu verwendender Phonemkatalog dort noch nicht vorliegt, diesen ganz oder teilweise zu übermitteln. Auch Änderungen oder Aktualisierungen von Phonemkatalogen können so der Sprachsyntheseeinrichtung übermittelt und damit verfügbar gemacht werden.

Eine weitere vorteilhafte Weiterbildung der Erfindung sieht vor, daß bei der Sprachsyntheseeinrichtung verschiedene, für die Phonemerzeugung relevante Parameter, wie z.B. die Länge, Lautstärke, Betonung, Stimmlage, Stimme etc. mit der Phoneme wiedergegeben bzw. ausgegeben werden, einstellbar sind; vorzugsweise für jedes einzelne Phonem. Die Wiedergabe der übermittelten Sprachdaten kann so unterschiedlichen Erfordernissen angepaßt werden. Beispielsweise kann die Wiedergabequalität i.a. verbessert werden, indem Laut- und/oder Stimmbesonderheiten, wie z.B. die Lautstärke und Tonhöhe von Silben oder ähnlicher Sprachelemente, vor dem Komprimieren aus den zu übermittelnden Sprachdaten extrahiert und der Sprachsyntheseeinrichtung zum Einstellen von Wiedergabeparametern übermittelt werden. Die Wiedergabeparameter können auch abhängig von einer Kennung eines die Sprachdatenen sendenden Teilnehmers eingestellt werden, um in die Sprachwiedergabe teilnehmerspezifische Elemente einfließen zu lassen, die z.B. den Teilnehmer anhand der wiedergegebenen Stimme identifizierbar machen können.

Weiterhin besteht die Möglichkeit, die von der Sprachdatenquelle zum Übermittlungsziel zu übermittelnden Sprachdaten als Kodezeichen in einem Speicher zwischenzuspeichern, um sie später ein oder mehrmals abzurufen. Durch die Reduzierung der Sprachdaten auf die Phoneme repräsentierenden Kodezeichen wird hierfür in der Regel wesentlich weniger Speichplatz benötigt als für ein Zwischenspeichern der unveränderten Sprachdaten.

Die Phoneme repräsentierenden Kodezeichen bilden darüber hinaus einen möglichen Ausgangspunkt für eine vielfältige Weiterverarbeitung der Sprachdaten. Beispielsweise können die Kodezeichen in den zugeordneten Phonemen entsprechende Buchstabengruppen umgesetzt werden und so die Sprachdaten als Text ausgegeben werden. Auch viele zeitgemäße Verfahren zur automatischen Spracherkennung beruhen auf einem Zerlegen der Sprache in Phoneme und können daher direkt mit dem erfindungsgemäßen Verfahren kombiniert werden.

Die Erfindung wird nachfolgend anhand eines bevorzugten Ausführungsbeispiels unter Bezugnahme auf die Figur näher erläutert.

Die Figur zeigt in schematischer Darstellung zwei über eine Übertragungsstrecke verbundene Kommunikationsanlagen.

In der Figur ist eine als Sender von Sprachdaten fungierende Kommunikationsanlage 1 und eine damit über eine Übertragungsstrecke US verbundene, als Empfänger der Sprachdaten fungierende Kommunikationsanlage 2 schematisch dargestellt. Die Kommunikationsanlage 1 enthält als Funktionskomponenten eine Sprachdatenquelle SD1 und ein daran angeschlossenes neuronales Netz NN mit Zugriff auf sprecherspezifische Phonemkataloge PN1 und PN2. In der Kommunikationsanlage 2 sind eine an sprecherspezifische Phonemkataloge PS1 und PS2 gekoppelte

Sprachsytheseeinrichtung SS und eine daran angeschlossene Sprachdatensenke SD2 dargestellt.

Die Beschränkung auf zwei sprecherspezifische Phonemkataloge je Kommunikationsanlage ist hier lediglich als beispielhaft anzusehen und dient der Vereinfachung der nachfolgenden Erläuterung.

In den sprecherspezifischen Phonemkatalogen PN1 und PN2 sind jeweils Phoneme hinterlegt, die durch Analyse repräsentativer Sprachdaten als charakteristische akustische Sprachelemente erkannt wurden. Die zur Erstellung der Phonemkataloge PN1, PN2 verwendeten repräsentativen Sprachdaten stammen dabei für jeden Phonemkatalog von einem spezifischen Sprecher. Den hinterlegten Phonemen ist jeweils ein für den betreffenden Phonemkatalog eindeutiges, abrufbares Kodezeichen zugeordnet.

Das neuronale Netz NN ist auf ein Erkennen der in den Phonemkatalogen PN1, PN2 hinterlegten Phoneme innerhalb eines Sprachdatenstroms trainiert, wobei einstellbar ist, welcher der Phonemkataloge PN1, PN2 bei dieser Phonemanalyse zugrundezulegen ist. Diese Einstellung kann beispielsweise dadurch erfolgen, daß dem neuronalen Netz eine dem jeweiligen Phonemkatalog PN1 bzw. PN2 entsprechende, sogenannte Gewichtsmatrix aufgeprägt wird.

Im Ausführungsbeispiel ist ein von der Sprachdatenquelle SD1 ausgehender Sprachdatenstrom von der Kommunikationsanlage 1 zur Sprachdatensenke SD2 in der Kommunikationsanlage 2 zu übermitteln, wobei die Sprachdaten vor der eigentlichen Übertragung zu komprimieren sind. Dazu wird von der Sprachdatenquelle SD1 zunächst eine Kennung eines die Sprachdaten sendenden Teilnehmers dem neuronalen Netz NN zugeführt, das abhängig von der übermittelten Kennung, einen diesem Teilnehmer zugeordneten sprecherspezifischen Phonemkatalog - hier PN1 - für die nachfolgende Phonemanalyse auswählt. Die Kennung wird daraufhin an die Kommunikationsanlage 2 weiterübermittelt.

Anschließend wird der zu übermittelnde Sprachdatenstrom dem neuronalen Netz NN zugeführt, das im Sprachdatenstrom auftretende Sprachdatenmuster nach den im ausgewählten Phonemkatalog PN1 hinterlegten Phonemen klassifiziert. Für jedes Sprachdatenmuster, das erfolgreich einem hinterlegten Phonem zugeordnet werden kann, wird anstelle des Sprachdatenmusters das dem hinterlegten Phonem im Phonemkatalog zugeordnete Kodezeichen über die Übertragungsstrecke US zur Kommunikationsanlage 2 übertragen. Sprachdatenmuster, die nicht erfolgreich einem hinterlegten Phonem zugeordnet werden können, werden - durch ein mit den Phonemen zugeordneten Kodezeichen nicht übereinstimmendes, spezielles Kodezeichen gekennzeichnet - unverändert übertragen.

In der Kommunikationsanlage 2 wird zunächst, abhängig von der übermittelten Kennung, ein der nachfolgenden Sprachsynthese zugrundezulegender Phomenkatalog - hier PS1 - ausgewählt. Der nachfolgend empfangene Strom von Kodezeichen wird sodann der Sprachsyntheseeinrichtung SS zugeführt, die für jedes Kodezeichen das diesem im ausgewählten Phonemkatalog PS1 zugeordnete Phonem erzeugt und der Sprachdatensenke SD2 zuleitet. Die durch das spezielle Kodezeichen gekennzeichneten, unverändert übertragenen Sprachdatenmuster werden direkt zur Sprachdatensenke SD2 weitergeleitet. Auf diese Weise wird der Sprachdatenstrom wieder annähernd rekonstruiert bzw. dekomprimiert.

Die Phonemkataloge PS1 und PS2 sind im Ausführungsbeispiel als konventionelle Speicher realisiert, in denen jedem enthaltenen Kodezeichen, eine zum Erzeugen des zugeordneten Phonems in der Sprachsyntheseeinrichtung SS erforderliche Information zugeordnet ist. Die Phonemkataloge PS1 und PS2 können durch Abrufen bzw. Auslesen der Phonemkataloge PN1 bzw. PN2 erstellt werden.

Patentansprüche

1. Verfahren zum Übermitteln von Sprachdaten von einer Sprachdatenquelle (SD1) zu einem Übermittlungsziel (SD2) unter Verwendung sprach- und/oder sprecherspezifischer Phonemkataloge (PN1, PN2, PS1, PS2), in welchen Phonemen entsprechende Sprachdatenmuster hinterlegt sind, und jedem Phonem jeweils ein eindeutiges Kodezeichen zugeordnet ist, mit folgenden Schritten:

a) abhängig von einer Kennung eines die Sprachdaten sendenden Teilnehmers wird ein diesem Teilnehmer zugeordneter sprecher- und/oder sprachspezifischer Phonemkatalog (PN1) ausgewählt,

b) die zu übermittelnden Sprachdaten werden einem auf das Erkennen der im ausgewählten Phonemkatalog (PN1) hinterlegten Phoneme trainierten neuronalen Netz (NN) zugeführt, das die Sprachdaten hinsichtlich der in diesen enthaltenen Phoneme analysiert,

c) für die in den Sprachdaten erkannten Phoneme wird das diesen im ausgewählten Phonemkatalog (PN1) jeweils zugeordnete Kodezeichen bestimmt,

d) die Kodezeichen werden zu einer Sprachsyntheseeinrichtung (SS) am Übermittlungsziel (SD2) übertragen,

e) der Strom empfangener Kodezeichen wird durch die Sprachsyntheseeinrichtung (SS) in eine Abfolge von den Kodezeichen in einem Phonemkatalog (PS1) jeweils zugeordneten Phonemen umgesetzt, und

f) diese Abfolge wird ausgegeben.

2. Verfahren zum Übermitteln von Sprachdaten von einer Sprachdatenquelle (SD1) zu einem Übermittlungsziel (SD2) unter Verwendung sprach- und/oder sprecherspezifischer Phonemkataloge (PN1, PN2, PS1, PS2), in welchen Phonemen entsprechende Sprachdatenmuster hinterlegt sind und jedem Phonem jeweils ein eindeutiges Kodezeichen zugeordnet ist, mit folgenden Schritten:

a) die zu übermittelnden Sprachdaten werden einem auf das

Erkennen verschiedener Sprachen und/oder Sprecher trainierten neuronalen Netz zugeführt, das eine Sprache, der die zu übermittelnden Sprachdaten angehören, und/oder einen Sprecher, von dem die zu übermittelnden Sprachdaten stammen, erkennt und ein Auswählen eines dieser Sprache und/oder diesem Sprecher zugeordneten, sprach- und/oder sprecherspezifischen Phonemkatalogs (PN1) veranlaßt,

b) die zu übermittelnden Sprachdaten werden einem auf das Erkennen der im ausgewählten Phonemkatalog (PN1) hinterlegten Phoneme trainierten neuronalen Netz (NN) zugeführt, das die Sprachdaten hinsichtlich der in diesen enthaltenen Phoneme analysiert,

c) für die in den Sprachdaten erkannten Phoneme wird das diesen im ausgewählten Phonemkatalog (PN1) jeweils zugeordnete Kodezeichen bestimmt,

d) die Kodezeichen werden zu einer Sprachsyntheseeinrichtung (SS) am Übermittlungsziel (SD2) übertragen,

e) der Strom empfangener Kodezeichen wird durch die Sprachsyntheseeinrichtung (SS) in eine Abfolge von den Kodezeichen in einem Phonemkatalog (PS1) jeweils zugeordneten Phonemen umgesetzt, und

f) diese Abfolge wird ausgegeben.

3. Verfahren nach Anspruch 1 oder 2,
dadurch gekennzeichnet,
daß die von der Sprachdatenquelle (SD1) zum Übermittlungsziel (SD2) zu übermittelnden Sprachdaten als Kodezeichen in einem Speicher zwischengespeichert werden.

4. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß Informationen über einen bei der Phonemanalyse der zu übermittelnden Sprachdaten benutzten Phonemkatalog (PN1) der Sprachsyntheseeinrichtung (SS) am Übermittlungsziel (SD2) übermittelt werden, um dort für eine Rekonstruktion der übermittelten Sprachdaten verwendet zu werden.

5. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß ein Phonemkatalog (PN1, PN2, PS1, PS2) während einer Sprachübermittlung unter Verwendung der zu übermittelnden Sprachdaten aktualisiert wird.

6. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß das neuronale Netz (NN) während einer Sprachübertragung auf das Erkennen der zu übermittelnden Sprachdaten trainiert wird.

7. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß der Strom der zu übertragenden Kodezeichen vor der Übertragung durch Anwendung eines weiteren Komprimierungsverfahrens komprimiert und am Übermittlungsziel (SD2) wieder dekomprimiert wird.

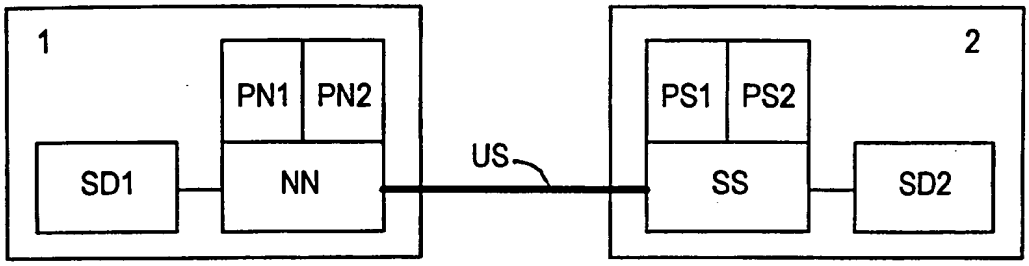
8. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß die Sprachsyntheseeinrichtung (SS) die Phoneme mit einstellbarer Länge, Lautstärke, Betonung, Stimmlage und/oder Stimme ausgibt.

9. Verfahren nach Anspruch 8,
dadurch gekennzeichnet,
daß Laut- und/oder Stimmbesonderheiten vor dem Komprimieren aus den zu übermittelnden Sprachdaten extrahiert und der Sprachsyntheseeinrichtung (SS) übermittelt werden, um dort zum Einstellen der Länge, Lautstärke, Betonung, Stimmlage und/oder Stimme mit der die Phoneme ausgegeben werden, verwendet zu werden.

10. Verfahren nach Anspruch 8 oder 9,
dadurch gekennzeichnet,
daß die Kennung des die Sprachdaten sendenden Teilnehmers

der Sprachsyntheseeinrichtung (SS) übermittelt wird, und dort die Länge, Lautstärke, Betonung, Stimmlage und/oder Stimme mit der die Phoneme ausgegeben werden, abhängig von der Kennung eingestellt wird.

FIG



INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/DE 98/03812

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G10L5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MAERAN O ET AL: "SPEECH RECOGNITION THROUGH PHONEME SEGMENTATION AND NEURAL CLASSIFICATION" IMTC-97: IEEE INSTRUMENTATION/MEASUREMENT TECHNOLOGY CONFERENCE, OTTAWA, CANADA, vol. 2, 19 - 21 May 1997, pages 1215-1220, XP000730875 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS see abstract see page 1219 - page 1220	1,3,4,7, 8
Y	--- -/--	2

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 April 1999	07/05/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Ramos Sánchez, U

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/DE 98/03812

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HAMPSHIRE II J B ET AL: "THE META-PI NETWORK: CONNECTIONIST RAPID ADAPTATION FOR HIGH-PERFORMANCE MULTI-SPEAKER PHONEME RECOGNITION" ICASSP'90: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, ALBUQUERQUE, USA, vol. 1, no. CONF. 15, 3 - 6 April 1990, pages 165-168, XP000146430 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS see abstract; figure 1 ---	2
A	EP 0 706 172 A (HUGHES AIRCRAFT CO) 10 April 1996 see abstract; figure 1 see column 3, line 13 - line 27 see column 6, line 35 - line 39 ---	1-4,7,8
A	WO 96 00962 A (LYBERG BERTIL ;TELIA AB (SE)) 11 January 1996 see abstract; figure 1 ---	1-4,7-9
A	EP 0 071 716 A (TEXAS INSTRUMENTS INC) 16 February 1983 cited in the application see abstract; figure 1 -----	1-3

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. ional Application No
PCT/DE 98/03812

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0706172 A	10-04-1996	JP 8194493 A	30-07-1996
		US 5832425 A	03-11-1998
WO 9600962 A	11-01-1996	SE 504177 C	02-12-1996
		EP 0767950 A	16-04-1997
		JP 10504404 T	28-04-1998
		SE 9402284 A	30-12-1995
		US 5694520 A	02-12-1997
EP 0071716 A	16-02-1983	US 4424415 A	03-01-1984
		US 4661915 A	28-04-1987
		JP 1869336 C	06-09-1994
		JP 58027200 A	17-02-1983

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PCT/DE 98/03812

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 6 G10L5/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 6 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	MAERAN O ET AL: "SPEECH RECOGNITION THROUGH PHONEME SEGMENTATION AND NEURAL CLASSIFICATION" IMTC-97: IEEE INSTRUMENTATION/MEASUREMENT TECHNOLOGY CONFERENCE, OTTAWA, CANADA, Bd. 2, 19. - 21. Mai 1997, Seiten 1215-1220, XP000730875 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS siehe Zusammenfassung siehe Seite 1219 - Seite 1220	1,3,4,7, 8
Y	--- -/--	2

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen.

[X] Siehe Anhang Patentfamilie

° Besondere Kategorien von angegebenen Veröffentlichungen :
"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
28. April 1999	07/05/1999

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Ramos Sánchez, U

1

Formblatt PCT/ISA/210 (Blatt 2) (Juli 1992)

INTERNATIONALER RECHERCHENBERICHT

Internales Aktenzeichen
PCT/DE 98/03812

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	HAMPSHIRE II J B ET AL: "THE META-PI NETWORK: CONNECTIONIST RAPID ADAPTATION FOR HIGH-PERFORMANCE MULTI-SPEAKER PHONEME RECOGNITION" ICASSP'90: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, ALBUQUERQUE, USA, Bd. 1, Nr. CONF. 15, 3. - 6. April 1990, Seiten 165-168, XP000146430 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS siehe Zusammenfassung; Abbildung 1 ---	2
A	EP 0 706 172 A (HUGHES AIRCRAFT CO) 10. April 1996 siehe Zusammenfassung; Abbildung 1 siehe Spalte 3, Zeile 13 - Zeile 27 siehe Spalte 6, Zeile 35 - Zeile 39 ---	1-4,7,8
A	WO 96 00962 A (LYBERG BERTIL ;TELIA AB (SE)) 11. Januar 1996 siehe Zusammenfassung; Abbildung 1 ---	1-4,7-9
A	EP 0 071 716 A (TEXAS INSTRUMENTS INC) 16. Februar 1983 in der Anmeldung erwähnt siehe Zusammenfassung; Abbildung 1 -----	1-3

1

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internales Aktenzeichen
PCT/DE 98/03812

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 0706172 A	10-04-1996	JP 8194493 A	30-07-1996
		US 5832425 A	03-11-1998
WO 9600962 A	11-01-1996	SE 504177 C	02-12-1996
		EP 0767950 A	16-04-1997
		JP 10504404 T	28-04-1998
		SE 9402284 A	30-12-1995
		US 5694520 A	02-12-1997
EP 0071716 A	16-02-1983	US 4424415 A	03-01-1984
		US 4661915 A	28-04-1987
		JP 1869336 C	06-09-1994
		JP 58027200 A	17-02-1983